MARKING SCHEME

MATHEMATICS Compulsory Part · Paper 1

Topic practice: Uses and Abuses of Statistics · 2027

This marking scheme has been prepared for markers' reference. It is not a set of model answers. Award method marks where the working is sound. Accept equivalent methods unless the question specifies otherwise.

Question	Solution	Marks	Remarks
SECTION A(1) (16 marks)			
1.	**22 boys and 18 girls** Stratified sampling allocates the sample in proportion to the stratum sizes. The sampling fraction is $\frac{40}{600} = \frac{1}{15}$. Number of boys $= 330 \times \frac{1}{15} = 22$. Number of girls $= 270 \times \frac{1}{15} = 18$. Check: 22 + 18 = 40.	3	
2.	**Self-selection bias (voluntary response bias). Only readers who are interested in the topic and have internet access are likely to respond, so the sample may over-represent certain groups and the results may not reflect the views of the whole population.** The survey relies on voluntary responses from magazine readers. This is a self-selection (voluntary response) sample. People who feel strongly about sport, or who have easy internet access, are more likely to take part. Hence the sample is not representative of the general population, leading to biased conclusions.	3	
3.	**(a) The truncated axis exaggerates differences between the bars. (b) Apparent ratio** $= \frac{20-10}{12-10} = 5$**, while true ratio** $= \frac{20}{12} = \frac{5}{3} \approx 1.67$**. The chart makes Dana's value appear 5 times Amy's, but it is actually only about 1.67 times.** (a) Because the vertical axis starts at 10, the visible height of a bar represents the value minus 10, not the value itself. Small differences in the actual values are magnified, so the chart can create a false impression of large differences. (b) On the truncated chart, the apparent height of Amy's bar is 12 − 10 = 2 units and that of Dana's bar is 20 − 10 = 10 units. Apparent ratio $= \frac{10}{2} = 5$. The true ratio is $\frac{20}{12} = \frac{5}{3} \approx 1.67$. Thus the chart exaggerates the comparison.	4	

Question	Solution	Marks	Remarks
4.	**(a) Population: all people in Hong Kong; sample: the 500 teenagers surveyed. (b) The sample consists only of teenagers, so it is not representative of the whole population of Hong Kong. (c) Use a random sample that includes people from all age groups, or conduct a stratified sample by age.** (a) The population is the entire group about which information is wanted: all people in Hong Kong. The sample is the subset actually surveyed: the 500 teenagers. (b) The sample is restricted to teenagers. Smartphone ownership among teenagers is likely to be higher than among young children or the elderly. Hence the sample is biased and cannot be used to estimate the proportion for the whole population. (c) A random sample drawn from all age groups, or a stratified sample with age as the stratum, would be more representative.	6	
SECTION A(2) (24 marks)			
5.	**(a) Production: 10, Packaging: 6, Administration: 4. (b) Number the 120 Production workers from 001 to 120, then use a random number table to choose 10 distinct numbers in that range. (c) Advantage: ensures each department is represented in proportion to its size. Disadvantage: requires a complete list of workers in each department and may be more time-consuming.** (a) Total workers $= 240$. Sampling fraction $= \frac{20}{240} = \frac{1}{12}$. Production: $120 \times \frac{1}{12} = 10$. Packaging: $72 \times \frac{1}{12} = 6$. Administration: $48 \times \frac{1}{12} = 4$. Check: $10 + 6 + 4 = 20$. (b) Assign each of the 120 Production workers a unique three-digit number from 001 to 120. Starting at a random point in a random number table, read three-digit numbers, ignoring numbers greater than 120 or already selected, until 10 distinct workers are chosen. (c) Advantage: the sample reflects the department structure of the workforce, reducing sampling error. Disadvantage: it requires a sampling frame listing all workers by department and may be administratively more complex than a simple random sample.	7	

TUTGATHER

Question	Solution	Marks	Remarks
6.	**(a) Convenience sampling. (b) It only includes employees who use the canteen on Monday morning, and the first 50 may not be representative; employees who arrive early or use the canteen may differ in transport habits. (c) Use simple random sampling from a complete list of all 1,200 employees, which gives every employee an equal chance of selection and reduces selection bias.** (a) The researcher selects the first available employees, so this is convenience sampling. (b) Bias may arise because: (i) only employees who visit the canteen on Monday morning are included, excluding those who bring lunch, work elsewhere, or are absent; (ii) the first 50 arrivals may be those who start early or live nearby, and their transport expenditure may differ systematically from others. (c) A simple random sample selected from the complete payroll list of 1,200 employees would give every employee an equal probability of selection. This removes the selection bias associated with choosing only canteen users at a particular time.	8	
7.	**(a) 25%. (b) The area ratio is** $(2.5/2)^2 = 1.5625$**, suggesting a 56.25% increase, which exaggerates the true 25% increase. (c) Use icons of the same width (or same base) and vary only the height in proportion to the data, or use a standard bar chart with a vertical axis starting at 0.** (a) Percentage increase $= \frac{100 - 80}{80} \times 100\% = 25\%$. (b) For similar 3-D icons, area scales as the square of the linear scale factor. The height ratio is $\frac{2.5}{2} = 1.25$, so the area ratio is $1.25^2 = 1.5625$. A reader comparing the areas would perceive an increase of about 56.25%, which is more than double the actual 25% increase. Thus the pictogram exaggerates the change. (c) To be fair, the icons should have the same width (or base) so that only the height varies in direct proportion to the number of visitors. Alternatively, a conventional bar chart with a vertical axis starting at zero should be used.	9	
SECTION B (10 marks)			

TUTGATHER

Question	Solution	Marks	Remarks
8.	**(a) 0.6. (b) The sample proportion is 0.6, but it is only an estimate; sampling error means the true proportion may differ, and a single sample of 800 cannot guarantee the exact population value. (c) A gym sample would over-represent regular exercisers, so the sample proportion would likely be higher than 0.6 and the conclusion would be invalid for the whole city. (d) Increase the sample size; limitation: it is more costly and time-consuming, and may still be biased if the sampling method is flawed.** (a) Sample proportion $= \frac{480}{800} = 0.6$. (b) The value 0.6 is a sample statistic, not the population proportion. Due to sampling error, different random samples would give different proportions. Although a sample of 800 is reasonably large, the true proportion could still be somewhat above or below 0.6. The newspaper states the figure as if it were an exact fact about all adults, which overgeneralises the result. (c) People at a gym are more likely to exercise regularly than the general adult population. This is a biased sample (selection bias). The sample proportion would probably be greater than 0.6, and any conclusion about the whole city would be invalid because the sample is not representative. (d) One way is to increase the sample size, which generally reduces the margin of error. Limitation: a larger sample is more expensive and time-consuming to collect, and increasing the size does not remove bias if the sampling method is not random.	10	

TUTGATHER